The Interplay of Pavlovian and Instrumental Processes in Devaluation Experiments: A Computational Embodied Neuroscience Model Tested with a Simulated Rat

Francesco Mannella **Marco Mirolli** **Gianluca Baldassarre**
Laboratory of Computational Embodied Neuroscience,
Istituto di Scienze e Tecnologie della Cognizione,
Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR),
Via San Martino della Battaglia 44, I-00185 Roma, Italy
{francesco.mannella, marco.mirolli, gianluca.baldassare}@istc.cnr.it

Abstract

This paper presents an embodied biologically-plausible model investigating the relationships existing between classical and instrumental conditioning. The architecture and functioning of the model is constrained with some some important anatomical and physiological assumptions drawn from the relevant neuroscientific literature. The model is validated by successfully reproducing the primary outcomes of some instrumental-conditioning devaluation tests conducted with normal and amygdala-lesioned rats. These experiments are particularly important as they show how the sensitivity to internal states (as satiety) exhibited by classical conditioning mechanisms can transfer to behaviors acquired on the basis of instrumental conditioning mechanisms. The results presented are relevant for both neuroscience and behavioural sciences as they are based on a model, constrained and validated at both neural and behavioural level, which indicates how internal states might modulate learning and performance of rigid habits so as to render to action some of the flexibility typical of goal-directed behaviour. The results are also relevant for autonomous robotics as they start to investigate, with an embodied system, how the use of sophisticated motivational systems might allow building robots capable of exhibiting some of the flexibility typical of organisms.

1 Introduction

The flexibility and capacity of adaptation of organisms greatly depends on their learning capabilities. For this reason, animal psychology has devoted great efforts to the study of learning processes. In particular, in the last century a huge body of empirical data have been collected around the two main experimental paradigms of 'classical conditioning' (Pavlov, 1927; Lieberman, 1993) and 'instrumental conditioning' (Thorndike, 1911; Skinner, 1938; Domjan, 2006).

Classical conditioning (also called 'Pavlovian conditioning') refers to an experimental paradigm in which a certain basic behaviour such as salivation or approaching (the 'unconditioned response' – UR), which is linked to a biologically salient stimulus such as food ingestion (the 'unconditioned stimulus' – US), gets associated to a neutral stimulus like the sound of a bell (the 'conditioned stimulus' – CS), after the neutral stimulus is repeatedly presented before the appearance of the salient stimulus. Such acquired associations are briefly referred to as 'CS-US' or 'CS-UR' associations (Pavlov, 1927; Lieberman, 1993).

Instrumental conditioning (also called 'operant conditioning') refers to an experimental paradigm in which an animal, given a certain stimulus/contexts such as a lever in a cage (the 'stimulus' – S), learns to produce a particular action such as pressing the lever (the 'response' – R), which produces a certain outcome such as the opening of the cage (the 'action outcome' – O), if this outcome is consistently accompanied by a reward such as the access to food. In this case, the acquired associations are briefly referred to as either 'S-R' associations, when the reactive nature of the acquired behaviour is stressed, or 'A-O' associations, when the goal-directed nature of behaviour

is stressed (Thorndike, 1911; Skinner, 1938; Domjan, 2006, see below).

This empirical work has been paralleled by the development, within the machine learning literature, of 'reinforcement learning algorithms' (Sutton and Barto, 1981, 1998), that is algorithms directed to provide machines with the capacity of learning new behaviors on the basis of rewarding stimuli (i.e. signals from the external environment that inform the machine about the achievement of desired goals). Interestingly, reinforcement learning algorithms have gained increasing interest within the empirical literature on animal learning as they represent theoretical models that can potentially furnish coherent explanations of organisms' learning processes. Indeed, a class of such models, the so-called *temporal-difference learning* algorithms (TD-learning), are currently considered as the best available theoretical accounts of several key empirical findings (Dayan and Balleine, 2002; Schultz, 2002; Houk et al., 1995).

Notwithstanding their success, standard reinforcement learning models suffer of several limitations from a biological point of view. In particular, three of the main drawbacks are as follows. First, such models ignore the role of internal states (e.g. hunger vs. satiety related to a certain type of food) in modulating the effects of 'external' rewards (e.g. the receival of such a food). Such kind of effects are exhibited by organisms, for example, in 'devaluation' experiments in which animals tend to change their reinforced behaviors in case the value of a rewarding stimulus, such as a food, is suddenly decreased through satiation or its association with poison. By ignoring the role of internal states in learning and behavior, current reinforcement learning models can not account for such effects.

Second, standard models do not take into account the important difference existing, within instrumentally acquired behaviors, between 'habits' and 'goal-direct actions', that is between those instrumentally acquired behaviors that are automatically triggered by *antecedent stimuli* and those that are controlled by their consequences (Yin and Knowlton, 2006). In fact, while classical (behaviorist) reinforcement learning theory assumed that all behaviors are elicited by some antecedent stimulus from the external environment, over the last few decades a significant body of research has demonstrated that animals are able to control their own behavior on the bases of the *expected outcomes* of their actions. The most important way to assess whether a behavior is driven by a stimulus or by an expected outcome is through a devaluation experiment: if a stimulus (S) elicits a response (R) only in case the associated reinforcing outcome (O) has not been devaluated, then it is evident that the behavior is driven by the outcome and not by the stimulus. Indeed, sensitiveness to the manipulation of outcome value has been proposed (cf. Yin and Knowlton, 2006) as part of the operational definition of *goal-directed behaviors* driven by action-outcome (A-O) associations as distinct from habits driven by stimulus-response (S-R) associations.

Third, standard models tend to conflate the notions of classical conditioning and instrumental conditioning. On the contrary, accumulating empirical evidence indicates that classical and instrumental conditioning are based on different processes that rely on distinct neural systems. Furthermore, such processes interplay in complex ways (Dayan and Balleine, 2002), as demonstrated, for example, by phenomena like 'PIT – Pavlovian-Instrumental Transfer' (where a conditioned stimulus that is predictive of reward can energize the execution of instrumentally acquired behaviours), and 'incentive learning' (where, under certain conditions, the valence of an action outcome needs to be re-learned to exert its effects on behaviour).

This paper starts to address these limitations by presenting a novel computational model which (a) is strongly rooted in the anatomy and physiology of the mammal brain, (b) is embodied in a simulated robotic rat, and (c) reproduces the results of empirical devaluation experiments conducted on both normal and amygdala-lesioned [1] rats (Balleine et al., 2003). Since, as indicated above, devaluation phenomena constitute the most evident demonstrations of both the role of internal states in modulating behavior and the distinction between habits and goal-driven behaviors, and since, as indicated below, here it is assumed that devaluation phenomena depend on the modulation of instrumental processes by Pavlovian processes, the attempt to reproduce these experiments constitutes the most appropriate way of addressing the aforementioned drawbacks of standard reinforcement learning models.

For these reasons, the proposed model constitutes the first working computational model that implements a coherent picture about the neural mechanisms underlying conditioning phenomena. More

[1] The amygdala, an almond-shaped group of nuclei located within each medial temporal lobe of the brain, is associated with a wide range of cognitive functions, including emotional regulation, learning, action selection, memory, attention and perception. Amygdala is involved in both aversive behaviours such as those involved in fear conditioning and taste aversion experiments (Blair et al., 2005; Knight et al., 2005; Maren, 2005), and appetitive behaviours (Baxter and Murray, 2002a; Cardinal et al., 2002; Balleine and Killcross, 2006). The functioning of amygdala relies on both its capacity to assign emotional valence to stimuli on the basis of input information related to internal body states, and on its capacity to associate neutral stimuli, biologically salient stimuli and innate responses. Some of its main functional sub-systems are (McDonald, 1998; Pitkänen et al., 2000) the 'central nucleus' (CeA), responsible for triggering innate responses and neuromodulation processes underlying learning and broad brain regulation, and the 'basolateral complex' (BLA), responsible for forming CS-US associations (Mannella et al., 2008).

specifically, the model is built upon the following three fundamental assumptions:

1. The *amygdala* constitutes the CS-US associator at the core of Pavlovian conditioning phenomena.
2. The *cortex-dorsolateral striatum* [2] pathway, forming S-R associations, constitutes the main actor involved in instrumental conditioning.
3. The *amygdala-nucleus accumbens* pathway 'bridges' classical conditioning processes happening in the amygdala and instrumental processes taking place in the basal ganglia, and so it allows producing the behaviours exhibited by rats in devaluation experiments.

Although the main goals of this papers have a scientific relevance, the research agenda that covers the work presented here is believed to posses the potential to produce useful outcomes also for technology. In fact, undoubtedly living organisms' behavior is characterized by a degree of autonomy and a flexibility that greatly overcomes those of current robots. A way to tackle these limits is to attempt to understand the mechanisms underlying organisms' behavioural flexibility as done here so as to use them in designing robot's controllers.

The rest of the paper is organised as follows. Sect. 2 illustrates the general methodological approach which guided the research reported in this paper. Sect. 3 reports the original experiments addressed by the model. Sect. 4 describes the simulated rats and environment used to tests the model. Sect. 5 contains a detailed description of the model. Sect. 6 reports the main results. Finally, Sect. 7 concludes the paper.

2 The Method Used: Computational Embodied Neuroscience

This paper addresses issues related to animal learning (in particular, conditioning phenomena) by following an approach which can be referred to as 'CEN – Computational Embodied Neuroscience' (cf. Prescott et al., 2003, 2006 which propose a research method which shares some principles with the approach proposed here). This method aims at providing general and sufficiently strong criteria for selecting models so as to produce *theoretical cumulativity* in the study of brain and behaviour. Indeed, the great amount of empirical data provided by neuroscience, psychology and the other related disciplines are seldom integrated by strong and general theoretical explanations, thus failing to produce a coherent picture of the phenomena under investigation. CEN aims to overcome these limits by relying upon the following principles:

1. *Evolutionary and adaptive framework.* The theory of evolution (Darwin, 1859) is the fundamental theoretical framework needed to understand biological phenomena and hence also to understand brain and behavior. This has at least three important implications. First, it tends to lead the focus of investigation on the *function* for the organisms' survival and reproduction of brain mechanisms and behavioural processes more than on their mere mechanical functioning. This is in contrast with some neuroscientific research which pose their attention only on neural mechanisms per-se, without trying to understand their role in organisms' adaptation. Second, it leads to recognise that much of the brain and behaviour functioning is related to the organisms' need to adapt their behaviour to varying environmental conditions during life and to avoid wholly encoding behaviour in the DNA. Third, it leads to recognize that the brain's architecture, physiology and plasticity mechanisms are the outcome of a 'blind' evolutionary process based on complex 'reproduction with variation' mechanisms. The first point leads to formulate scientific questions with an emphasis on function. The last two points lead to investigate (and to explicitly model, see below) the learning and evolutionary processes that 'historically' lead organisms' brain and behaviour to be as they are (Parisi et al., 1990; Zlatev and Balkenius, 2001; Weng et al., 2001).
2. *Complex systems framework.* The brain and the brain-body-environment set are complex systems. Complex systems are systems formed by many parts (e.g. brain neurons) which interact via local rules (e.g. activation potentials). These interactions lead to the emergence of the global behaviour of the system without (fully) relying upon centralised coordination mechanisms, but on self-organisation principles such as positive feedback and negative feedback (Camazine et al., 2001; Baldassarre, 2008). This not only implies that brain and behaviour have to be studied with computational models (see below) but also that their understanding has to rely upon the concepts of complex-system theory.
3. *Computational models.* The investigation of brain and behaviour conducted on the basis of

[2] The striatum is the input portion of the basal ganglia, a set of forebrain subcortical nuclei that are traditionally considered to be responsible for instrumental conditioning phenomena (i.e. the locus of S-R associations: Packard and Knowlton, 2002; Yin and Knowlton, 2006). In rats, the striatum can be divided in: (a) dorsolateral striatum, mainly underlying motor-execution functions; (b) dorsomedial striatum, playing a role in motor-preparation and cognitive functions; (c) nucleus accumbens (or ventral striatum), considered an important interface (Mogenson et al., 1980) between the motivational processes taking place in the limbic system and the motor processes taking place within the rest of the basal ganglia and cortex)

empirical experiments and observations (such as those of neuroscience, psychology and ethology) should be accompanied by the instantiation of theories into formal computational models, that is computer programs that simulate the mechanisms underlying brain processes and produce behaviour as an emergent outcome of their functioning. The rationale behind this principle is that the brain, and the brain-body-environment set, are complex systems, and theories expressed only through words or analytically-solved equation systems can give only limited, not very generative accounts of these phenomena.

4. *Constraints from behaviour.* The computational models used to instantiate the theories have to be capable of reproducing the investigated behaviour, in line with what is proposed by 'artificial ethology' (Holland and McFarland, 2001). Furthermore, the comparison between the model and the target behaviour should be done on a detailed basis (i.e., with reference to the outcomes of specific target experiments) and possibly in quantitative terms (i.e., not just with vague, qualitative comparisons).

5. *Constraints from brain.* Challenging models with the request to account for specific behavioural data, is not enough as, given a behaviour, many alternative models capable of reproducing it can always be built. For this reason, a second fundamental source of constraints for models are the data on the anatomy and physiology of the brain. These data should be used in two ways. First, for choosing the assumptions that drive the design of the architecture, functioning, and learning mechanisms of the models. Second, for testing the low-level predictions produced by the models (i.e. the predictions produced at the neural level). This principle comes from computational neuroscience (Sejnowski et al., 1988) urging computational models to keenly account for data on brain.

6. *Embodiment.* In line with the ideas proposed by the 'animats approach' (Meyer and Wilson, 1991) and 'artificial life' (Langton, 1987), models should be capable of reproducing the addressed behaviours within 'whole' autonomous systems acting on the basis of circular interactions with the environment mediated by the body (the sensors, the actuators and the internal body states).

   Indeed, the brain generates behaviour by forming a large dynamical complex system together with the body and the environment (Clark, 1997; Nolfi and Floreano, 2000; Nolfi, 2006), so a full understanding of brain and behaviour needs to rely on models that take into consideration this fundamental fact.

   The principle has two implications. First, the computational models should involve the simulation of both organisms' brains and their body and environment, thus letting behaviors emerge from the interactions between those systems. Second, models should aim at being *scalable to realistic setups*, that is they should be capable of functioning with realistic sensors (e.g. retinas), realistic actuators (e.g. bodies should be governed by realistic Newtonian dynamics), realistic scenarios (e.g. objects and supports with complex textures, shapes, and dynamics), and noise (affecting sensors, actuators, environment, etc.).

7. *Generality.* Computational models should aim to reproduce and account for an increasing amount of data from an increasing number of different experiments. This principle is important as it is a strong drive towards the production of comprehensive accounts and general theories of brain and behaviour, against the tendency to generate many unrelated and mutually incompatible theories each accounting for only a limited set of empirical data. This principle is in line with the 'spirit' of both 'systems computational neuroscience' (Brody et al., 2004), that aims at explaining the functioning of whole brain systems rather than specific areas or physiological/chemical mechanisms, and with the computationally-informed approaches proposed within psychology (Newell, 1973), which stress the need of identifying general principles and theories on behaviour.

3 Target Experiment

The target data addressed with the model are reported in Balleine et al. (2003) which illustrates various experiments directed to investigate the relations existing between the manipulation of the value of primary rewards (devaluation) and instrumental conditioning, and the role that Amygdala (Amg) plays in them. The present work focusses on 'Experiment 1' reported in the article, a standard 'devaluation test'. This experiment is particularly well-suited to test four important processes that the model aims to capture: (a) the association between neutral stimuli (e.g. the sight of a lever) and biologically salient stimuli (e.g. food); (b) the dependence of the evaluation of external stimuli (e.g. food, levers, etc.) on internal states; (c) learning and use of 'habits', that is stereotyped and rigid behaviours acquired during prolonged periods of trial-and-error learning (these processes are well-captured by standard reinforcement-learning models); (d) the influence of internal states, mediated by the associations between neutral-stimuli and biologically-salient-stimuli, on the selection and triggering of habits.

In two preliminary phases of the experiment, eight sham-lesioned plus eight rats whose basolateral amygdala complex (BLA) was lesioned were trained in *separate* trials to press a lever or pull a chain to obtain two different kinds of food, respectively Noyes pellets and maltodextrin. The training phase was followed by an extinction test lasting 20 mins (divided in groups of 2 mins) where: (a) *both* manipulanda were present in the experimental chamber; (b) half of the rats had been previously satiated with Noyes pellets while the other half with maltodextrin. The main result indicated that during the first two minutes of the test non-lesioned rats performed the action corresponding to the manipulandum of the non-satiated food with a much higher rate with respect to the other manipulandum. On the other hand, BLA-lesioned rats did not show any devaluation effect: they performed the two actions at the same rate. These experiments clearly demonstrate that BLA plays a fundamental role in the transfer of the diminished hedonic value of food to instrumentally acquired habits. As we shall see in Sect. 5, this key finding is central for clarifying the relationship existing between Pavlovian and instrumental conditioning.

4 The Simulated Organism, Environment and Experiments

In line with the CEN approach (Sect. 2), the model presented here was tested within an embodied system. Although we are aware that the role of the 'degree of embodiment and situatedness' of the model and simulations presented here is rather limited (e.g. the sensors and actuators used are rather simplified, low-level behaviors are hardwired, etc.), nevertheless the use of a simulated organism and experimental setup forced us to design a model potentially capable to cope with the difficulties posed by more realistic setups. For example, the noise of behaviour execution, and the randomly variable duration of the trials, actions' execution, and rewarding effects posed interesting challenges to the robustness of the learning algorithms of the model.

The model was tested with a simulated robotic rat ('ICEAsim') developed within the EU project ICEA on the basis of the physics 3D simulator Webots^{TM}. The model was written in Matlab^{TM} and was interfaced with ICEAsim through a TCP/IP connection. The robotic setup used to test the model is shown in Fig. 1 and it is now briefly described skipping details not central for the paper's goals. The training and test environment is composed by a grey walled chamber containing a yellow lever, a red chain, and a grey food-dispenser that turns green or blue when respectively food A or food B is delivered in it. When 'pressed' or 'pulled', the lever and chain make respectively food A or B (the rewarding stimuli) available at the dispenser.

The simulated rat is a wheel-chair robot equipped with various sensors. Among these, the experiments reported here use two cameras (furnishing a panoramic 300 degrees view) and the whisker sensors. The rat uses the cameras to detect the lever, the chain and the food dispenser, in particular their presence/absence (via their color) and their (egocentric) direction. The rat uses the whiskers, activated with one if bent beyond a certain threshold and zero otherwise, to detect contacts with obstacles. The rat is also endowed with *internal* sensors related to satiety for either food A or B (these sensors assume the value of one when the rat is satiated, and zero otherwise). The rat's actuators are two motors that can independently control the speed of the two wheels.

For simplicity, the information fed to the model is only related to the presence/absence of the lever and chain in the test chamber and food A and food B in mouth, whereas the other information is used to control a four low-level hardwired behavioral routines. These routines, triggered either by the model or directly by stimuli, are as follows: (1) *obstacle avoidance routine*: this routine, triggered by the whiskers, 'overwrites' all other actions to avoid obstacles; (2, 3) *lever press routine* and *chain pull routine*: these routines, activated by the model, cause the rat to approach the lever/chain on the basis of their visually detected direction; when the lever/chain are touched they activate the food delivery in the dispenser; (4) *consummatory routine*: when the dispenser turns green or blue (this signals the presence of food in it), the rat approaches and touches it ('consummation' of the food) so causing the perception of respectively food A or food B in mouth; the routine ends after the rat touches the dispenser ten times.

The simulated devaluation experiment is divided in a training phase and two test phases. The training phase lasts 8 mins and the two test phases 2 mins each. Each phase is divided in trials that end either when the rat executes the correct action and consumes the food or after a $15s$ timeout (the duration of the experimental phases is shorter than the duration of the original experiment's phases as the limited complexity and number of available actions of the simulated rat allowed a faster learning). In each trial the rat is set in the middle of the chamber with an orientation randomly set between the lever and the chain direction. In the trials of the training-phase either the lever and food A or the chain and food B are used in an alternate fashion and the rat is always 'hungry' (the two satiation sensors are set to 0). In the two test phases, the rat is respectively satiated either with food A (the satiation sensors for food A and B are respectively set to one and zero) or with food B. In all trials of the two test phases

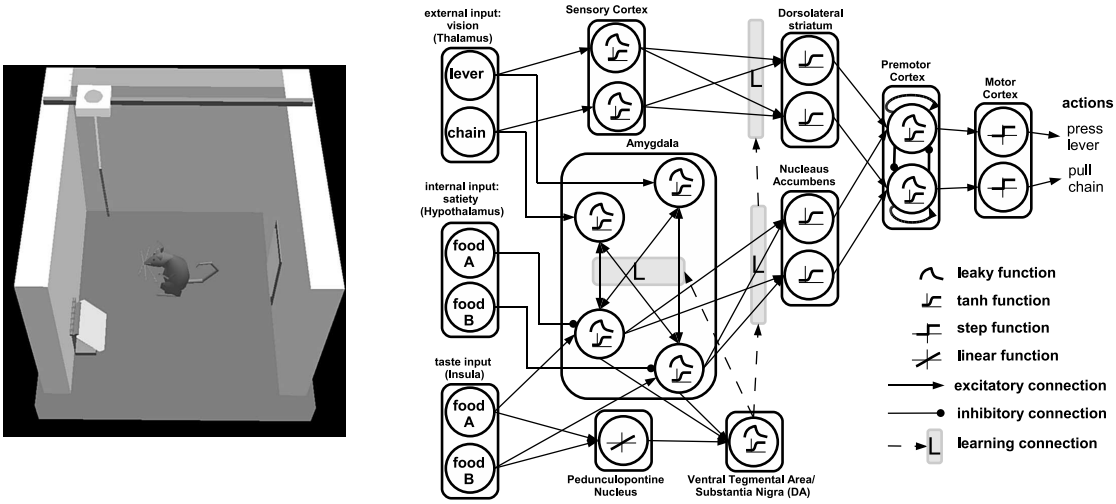

Figure 1: Left: A snapshot of the simulator, showing the simulated rat at the center of the experimental chamber, the food dispenser (behind the rat), the lever (at the rat's left hand side) and the chain (at the rat's right hand side). Right: The architecture of the model.

both manipulanda are present and the rat is evaluated in extinction (i.e. without delivery of food). The experiment (the three phases) was run 20 times with 'unlesioned' artificial rats and 20 times with 'lesioned' rats.

5 The Model

The model (Fig. 1) is formed by three major components: (a) the *amygdala* which instantiates an CS-US associator and is at the core of Pavlovian conditioning (Baxter and Murray, 2002b; Cardinal et al., 2002); (b) the *cortex-dorsolateral striatum* pathway which learns, via instrumental conditioning processes, habits based on S-R associations (Yin and Knowlton, 2006); (c) the *amygdala-nucleus accumbens* pathway which 'bridges' classical conditioning processes happening in the amygdala and instrumental processes taking place in the basal ganglia (Baxter and Murray, 2002b; Cardinal et al., 2002).

The model's input is formed by six neurons activated by the sensors illustrated in Sect. 4: two neurons encode the presence/absence of the lever and the chain (s_{lev} and s_{cha}), two neurons encode the presence/absence of food A food B in the rat's mouth (s_{fA} and s_{fB}), and two neurons encode the satiation for food A and food B (s_{sfA} and s_{sfB}).

5.1 The Amygdala: A CS-US Associator

The associator implements Pavlovian conditioning through the association between CSs and USs ('stimulus substitution'). In real brains this role seems to be played by the Amg (Baxter and Murray, 2002b; Cardinal et al., 2002). There are massive reciprocal connections between the Amg and several brain areas, including: inferotemporal cortex (IT), insular cortex (IC), prefrontal cortex (PFC), and hippocampus (Hip) (Price, 2003; Rolls, 2005; Baxter and Murray, 2002b; Cardinal et al., 2002). Furthermore, Amg receives inputs from posterior intralaminar nucleui of thalamus (PIL) (Shi and Davis, 1999). These connections underlie an interplay between processes related to perceived (or represented) external context (IT, PFC, Hip) and processes related to internal states (IC, PIL). In general, Amg can be seen as playing the function of assigning a subjective valence to external events on the basis of the animal's internal context (needs, motivations, etc.), and to use this to both regulate learning processes and directly influence behavior.

The model's associator, which is considered as an abstraction of the processes taking place in the Amg, performs 'asynchronous learning/syncronous functioning' associations. First, stimuli perceived in different times are associated (CSs are associated to USs): this associative learning takes place if USs cause a dopamine (DA) release (see below). When the association is established, CSs are able to synchronously re-activate the USs' representations in Amg. The associator is composed by a vector $\mathbf{amg} = (amg_{lev}, amg_{cha}, amg_{fA}, amg_{fB})'$ of four reciprocally-connected leaky neurons that process the input signals as follows:

$$\tau_{\mathbf{amg}} \cdot \dot{\mathbf{amg}}_p = -\mathbf{amg}_p + (s_{lev},\ s_{cha},\ (s_{fA} - s_{sfA}),\ (s_{fB} - s_{sfB}))' + \mathbf{W}_{amg} \cdot \mathbf{amg} \quad (1)$$

$$\mathbf{amg} = \varphi[tanh[\mathbf{amg}_p]]$$

where $\mathbf{amg}_p$ are the activation potentials of Amg, $\varphi[.]$ is a positive linear function ($\varphi[x] = 0$ if $x \leq 0$

and $\varphi[x] = x$ otherwise) and $\mathbf{W}_{amg}$ is the matrix of all-to-all lateral connection weights within Amg. Note that while external stimuli have a binary representation (0/1 for absence/presence), internal stimuli *modulate* the representation of external stimuli. In particular s_{sfB} and s_{sfB} assume a value in $\{0, 5\}$ when the corresponding satiation has respectively a low or high value, and this simulates the fact that satiation for a food inhibits the hedonic representation of such food within Amg. This assumption is supported by evidence indicating that a similar computation is performed in the secondary taste areas of the prefrontal/insular cortex (Rolls, 2005) connected with Amg. This part of the model is particularly important because, as we shall see, it mediates the influence of the shifts of primary motivations on both learning and behavior.

The associator's learning is based on the *onset* of input signals, detected as follows. First, 'leaky traces' **tr** of the derivatives of **amg**, trunked to positive values, are computed:

$$\tau_{\mathbf{tr}} \cdot \dot{\mathbf{tr}} = -\mathbf{tr} + C_{Amg} \cdot \varphi[\dot{\mathbf{amg}}] \quad (2)$$

where C_{Amg} is an amplification coefficient. Second, the derivatives of **tr** are computed: when positive, these derivatives detect the onset of the original signals, whereas when negative they detect the fact that some time elapsed since such onset took place.

The weights between Amg's neurons are updated on the basis of the signs of $\dot{\mathbf{tr}}$ and the DA signal (see below). In particular, when (and only when) the derivative of the presynaptic neuron's trace is negative and the derivative of the postsynaptic neuron's trace is positive (notice that this happens when the presynaptic neuron fires *before* the postsynaptic neuron) the related connection is strengthened (for all couples of neurons this condition is encoded in the matrix **L** with Boolean elements):

$$\Delta\mathbf{W}_{amg} = \eta_{amg} \cdot \varphi[da - th_{da}] \cdot \mathbf{L} \quad (3)$$

where η_{amg} is a learning rate coefficient, da is the dopamine signal and th_{da} is a threshold over which dopamine elicits learning.

DA release (corresponding to activation in the ventral tegmental area, VTA, and in the substantia nigra pars compacta, SNpc) is triggered by Amg through the units representing the hedonic impact of food and by the primary reward signals received from the peduncolo pontine tegmental nucleus (PPT) (Kobayashi and Okada, 2007):

$$\begin{aligned}\tau_{da_p} \cdot \dot{da}_p = &-da_p + da_{baseline} + \\ &w_{amg-da} \cdot (amg_{f_A} + amg_{f_B}) + \\ &w_{ppt-da} \cdot ppt\end{aligned} \quad (4)$$

$$da = \varphi[tanh[da_p]]$$

where $ppt = s_{fA} + s_{fB}$ is the PPT's primary reward signal. DA drives learning in both the associator and the action selectors (see Sect. 5.2 and 5.3).

5.2 The Cortex-Dorsolateral Striatum Pathway: A S-R Associator

The action selector based on the cortex-dorsolateral striatum learns 'habits' (S-R associations) through reinforcement learning processes. In real brains this function might be implemented in the cortex-dorsolateral striatum pathway (Yin and Knowlton, 2006). In the model this component receives s_{lev} and s_{cha} as input and, on the basis of this, selects one of the two lever-press/chain-pull actions (together with NAc, see Sect. 5.3).

The component is formed by four layers of neurons corresponding to four vectors: (a) a visual sensory cortex (SC) leaky-neuron layer: **sc**; (b) a neuron layer corresponding to DLS's encoding of the 'votes' for the two actions: **dls**; (c) a neuron layer corresponding to premotor cortex (PM), formed by reciprocally inhibiting neurons that implement a competition for selecting one of the two actions (this function might be implemented by the reciprocal thalamo-cortical connections, Dayan and Balleine, 2002): **pm**; (d) a layer corresponding to motor cortex (M), representing the selected action with a binary code: **m**.

The visual leaky-neuron layer processes the input signal in a straightforward fashion:

$$\tau_{\mathbf{sc}} \cdot \dot{\mathbf{sc}}_p = -\mathbf{sc}_p + (s_{lev},\ s_{cha})' \quad (5)$$

$$\mathbf{sc} = \varphi[tanh[\mathbf{sc}_p]]$$

SC is fully connected with DLS. DLS's (non-leaky) neurons collect the signals from SC that tend to represent the evidence ('votes') in favor of the selection of either one of the two actions:

$$\mathbf{dls}_p = \mathbf{W}_{(sc-dls)} \cdot \mathbf{sc} \quad (6)$$

$$\mathbf{dls} = \varphi[tanh[\mathbf{dls}_p + dls_{baseline}]]$$

The selection of actions is performed on the basis of these votes (and NAc's votes, see Sect. 5.3) through a competition taking place between the leaky neurons of DLS:

$$\begin{aligned}\tau_{pm} \cdot \dot{\mathbf{pm}}_p = &-\mathbf{pm}_p + w_{dls-nac-pm} \cdot \\ &(\mathbf{dls} + \mathbf{nac}) + \mathbf{W}_{pm} \cdot \mathbf{pm} + \mathbf{n}\end{aligned} \quad (7)$$

$$\mathbf{pm} = \varphi[tanh[\mathbf{pm}_p]]$$

where $w_{dls-nac-pm}$ is a coefficient scaling the votes, $\mathbf{W}_{pm}$ are the PM's lateral connection weights, and **n**

is a noise vector with components uniformly drawn in $[-n, n]$. The assumption for which the action selection takes place within PM, used here for simplicity, raises interesting complex problems which are discussed in Sect. 7.

When one of the **pm** neurons reaches an activation threshold th_A, the execution of the corresponding action is triggered via M :

$$\mathbf{m} = \psi[\mathbf{pm} - th_A] \tag{8}$$

where $\psi[x]$ is a step function ($\psi[x] = 0$ if $x \leq 0$ and $\psi[x] = 1$ otherwise). Once the execution of the routine corresponding to the selected action terminates, the connection weights between SC and DLS, $\mathbf{W}_{sc-dls}$, are modified according to the dopamine signal (this might be null in the case the wrong action has been selected):

$$\Delta\mathbf{W}_{sc-dls} = \eta_{sc-dls} \cdot \varphi[da - th_{da}] \cdot \mathbf{m} \cdot \mathbf{sc}' \tag{9}$$

where η_{sc-dls} is a learning coefficient. Note that here M activations were directly used to train both the DLS and NAc (see Sect. 7 on this strong assumption).

5.3 The Amygdala-Nucleus Accumbens Pathway: A Bridge between Pavlovian and Instrumental Processes

The Amg-NAc pathway 'bridges' Pavlovian processes to instrumental processes in that it learns A-O associations between the USs encoded in Amg (which might be thought of as desired outcomes of actions, corresponding to ingested elements of food in the presence of hunger for such food, elicited by the CSs, e.g. the sight of a lever) and actions encoded in the SC-DLS-PM pathway. In real brains this function might be implemented by the neural pathway connecting the BLA nuclei of Amg to NAc (Baxter and Murray, 2002b). In the model the pathway is implemented through an all-to-all connection matrix $\mathbf{W}_{amg-nac}$ linking the Amg's hedonic representation of food, amg_{fA} and amg_{fB}, to the NAc's (non-leaky) neurons:

$$\mathbf{nac}_p = \mathbf{W}_{amg-nac} \cdot (mg_{fA},\ amg_{fB})' \tag{10}$$

$$\mathbf{nac} = \varphi[tanh[\mathbf{nac}_p + nac_{baseline}]]$$

NAc's neurons play the same function as DLS neurons, namely they represents 'votes' that bias the action competition taking place in PM. Similarly to SC-DLS connections, Amg-NAc connections $\mathbf{W}_{amg-nac}$ are modified, after action execution, on the basis of the dopamine signal:

$$\Delta\mathbf{W}_{amg-nac} = \eta_{amg-nac} \cdot\ \varphi[da - th_{da}] \cdot \mathbf{m} \cdot (amg_{fA},\ amg_{fB}) \tag{11}$$

where $\eta_{(amg-nac)}$ is the learning rate coefficient. Note that in the experiments reported in Sect. 6 the lesions of rats' BLA have been simulated by setting the Amg-NAc connections $\mathbf{W}_{amg-nac}$ to zero.

The importance of the Amg-NAc action selector resides in the fact that its 'votes' for the various actions can be *modulated on the fly* by the system's motivational states, e.g. by satiety for either one of the two foods. In general, this mechanisms opens' up the possibility for the motivational-sensitive Pavlovian system (mainly the Amg in the model) to exert a direct effect on actions without the need of relearning processes, as it will be exemplified by the devaluation experiments illustrated in the next section.

5.4 Parameters' Setting and Justification

The model's parameters were set as follows. The model's equations were integrated with a $\Delta t = 50ms$ time step: this rather long value allows running fast simulations and at the same time avoiding stability problems. The decay coefficient of most leaky neurons of the model were set to a rather high value (which implies a slow dynamics) as such neurons are intended to abstract the activation of *populations* of real neurons: $\tau_{sc} = 500ms$, $\tau_{amg} = 500ms$, $\tau_{pm} = 500ms$. The decay of DA is set to a rather low value (which implies a fast dynamics) to reproduce the fast dynamics of phasic dopamine bursts underlying learning (Schultz, 2002): $\tau_{da} = 50ms$. The decay of learning traces is set to a high value in order to allow the association between stimuli having onsets separated by time intervals ranging within few seconds, as it happens in real rats: $\tau_{tr} = 1000ms$. The trace-derivative amplification coefficient is set to a high value to suitably amplify the low value of the derivative of Amg neurons' activation: $C_{amg} = 50$. The NAc and DLS baseline coefficients, and the weights connecting them to PM, are set to suitable values so as to not overcome the action triggering threshold in PM: $nac_{baseline} = .3$, $dls_{baseline} = .3$, $w_{dls-nac-pm} = .5$, $th_A = .6$. The baseline of DA is set below the DA threshold which trigger learning: $th_{DA} = .6$, $da_{baseline} = .3$. The Amg-DA connections are set to a value lower than the PPN-DA to have a DA signal stronger for primary rewards (US) than for secondary rewards (CS): $w_{amg-da} = .3$, $w_{ppn-da} = .6$. On the other side, the noise level is set to a rather high value to allow triggering of actions in the initial exploratory phase where the signal activation of NAc and DLS is null or low: $n = .6$. Learning coefficients are set to relatively low values to have a progressive stable learning: $\eta_{amg} = .015$, $\eta_{amg-nac} = .02$, $\eta_{sc-dls} = .02$. The weights of lateral connections between PM neurons are set to values which lead to a stable and reliable competition:

$$w_{pm} = \begin{pmatrix} 1 & -.5 \\ -.5 & 1 \end{pmatrix}.$$

6 Results

This section describes the basic functioning of the model on the basis of Fig. 2. The figure shows the activations of various neurons related to the lever (data related to the chain are omitted as qualitatively similar) during both the training and testing phases of an experiment run with a non-lesioned simulated rat. It also shows the activations of the same neurons in the two test phases of a lesioned rat.

At the beginning of the training phase, the baseline activations of DLS and NAc (dls_{lev}, nac_{lev}), together with noise, are sufficient to occasionally trigger the execution of an action (m_{lev}) by the competition taking place in PM (pm_{lev}). When the behavioral routine corresponding to the selected action is appropriate for the environment configuration (e.g. 'lever press' in the presence of lever), the dispenser becomes green and the rat approaches it and consumes the corresponding food (s_{fA}). The food consumption activates the internal hedonic representation of food in Amg (amg_{fA}) and hence the neurons in VTA/SNpc release DA in DLS. This drives the learning of the cortex-dorsolateral striatum instrumental pathway. The effect of these events is that after a few learning trials the model learns the S-R habits which perform the action which is appropriate to the current context: 'sight of lever-press lever' and 'sight of chain-pull chain'. The progress of habits' learning can be seen in terms of: (a) the increase of DLS's votes for the press lever action (dls_{lev}) in the trials in which the lever is present (s_{lev}); (b) the increase of the regularity of the peaks of the food A amygdala neurons (amg_{fA}); (c) the DA release in VTA/SNpc (da).

When instrumental S-R associations begin to form, the vision of the neutral stimuli of the lever (s_{lev}, amg_{lev}) starts to be reliably followed, within a relatively small time interval, by the food perception in mouth (s_{fA}). The food perception, as mentioned above, causes a DA release (da). This contingency and the DA signal allow the Pavlovian learning taking place within Amg to 'take off' and form CS-US associations between the lever and Amg's food A representation. This is evident from the fact that after a few successful trials the Amg's food-A neuron (amg_{fA}) not only shows an activation peak when food A is delivered but it is also pre-activated by the presence of the lever: this reveals that a Pavlovian association is being acquired between the CS (lever) and the US (food). Note how these processes show a rather interesting interaction between Pavlovian and instrumental processes. In the model, as in organisms (Lieberman, 1993), Pavlovian CS-US associations can form only if the two stimuli are separated by a time lag lasting at maximum few seconds (in the model, this is due to the dynamics of Amg's traces, see Eq. 2 and Sect. 5.4). As with the progress of the S-R instrumental learning, e.g. involving the 'sight of lever-press lever' association, the sight of the lever (CS) is followed progressively more readily and regularly by the food (US), this allows Pavlovian processes to form the association CS-US, which would not otherwise form (roughly speaking, it might be said that 'Pavlov' observes and registers a contingency suddenly appearing in the environment due to 'Skinner').

The pre-activation of the amg_{fA} neuron due to the perception of the conditioned stimulus is responsible for the early DA release (da) which anticipates the future delivery of reward. Even if this process does not play any particular function in the current model, it reproduces an important well-known phenomenon observed in real animals (Schultz, 2002), and shows how Amg can play an important role in the neuromodulation of brain, in this case the DA release.

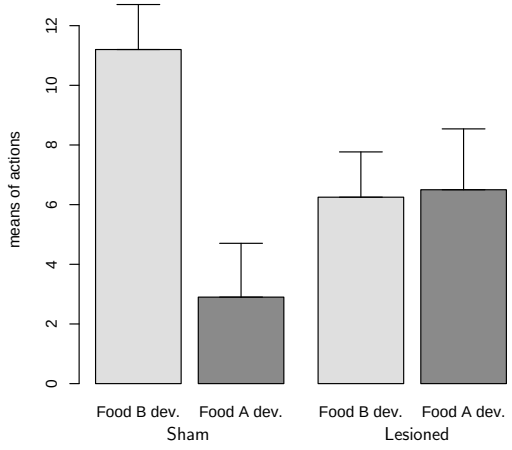

Figure 3: Averages and standard deviations of the number of actions selected (y-axis) by Sham and BLA-lesioned rats during different tests involving devaluation of either Food A or Food B.

A last important learning process takes place in the Amg-NAc pathway. This process is at the basis of the influence of Amg on the selection of habits taking place in the SC-DLS pathway. Once the CS-US associations are formed in the Amg, CSs, such as the lever, can trigger the activation of the Amg's hedonic representation of the related food and, via this, influence DLS action selection via NAc. This process is shown by the fact that, after some training, NAc starts to activate and to vote for the correct actions (nac_{lev}). The importance of the formation of this Stimuli-Amg-NAc-PM pathway resides in the fact that it constitutes the fundamental bridge between the the Pavlovian processes happening in the Amg

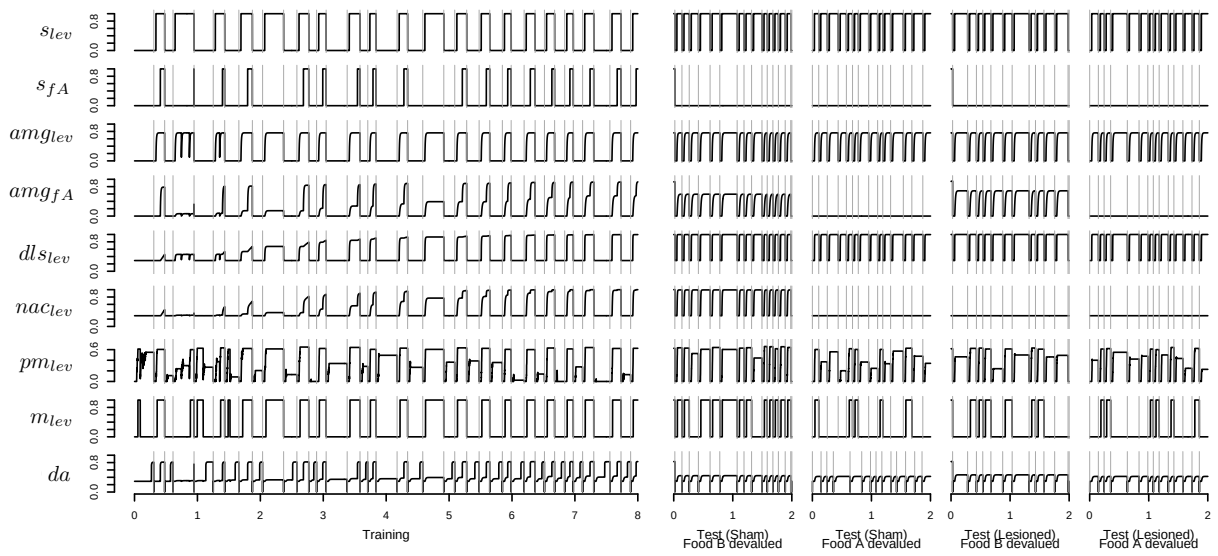

Figure 2: Activations of some key neurons of a non-lesioned rat during the a training phase (first block); activations of the same neurons in two test phases where the same rat was satiated either with food A or B (second and third block); activations of the same neurons of a BLA-lesioned rat in two similar test phases (fourth and fifth block). Trials are separated by short vertical lines.

and the instrumental processes happening in the SC-DLS pathway. We claim that this pathway plays a central role in the flexibility demonstrated by real organisms. In particular, it is through this pathway that instant motivational manipulations that characterize Pavlovian conditioning are able to directly affect instrumentally learned behaviors, as in the devaluation tests which are now illustrated.

Let us now focus on the test with sham rats (Fig. 2, second and third block). During the two test phases, in which the rat perceives both the lever and the chain, the satiety of respectively food A or B are kept at its maximum level, namely five (the other satiety level is kept at zero). Recall that the tests are performed 'in extinction', that is without food delivery (see s_{fA}). The satiety for a food causes a strong inhibition to the Amg's hedonic representation of such food. As a consequence both the direct consumption of that food and the perception of the conditioned stimulus previously associated with it cannot elicit the related Amg's hedonic reaction. This is shown by the lack of activation of the Amg's Food-A neuron (amg_{fA}) during the second test phase when the rat is satiated with food A.

The perception of both the lever and the chain during the tests leads DLS to 'vote' for both the lever press and chain pull actions at the same time (compare the dls_{lev} activation in the two test phases). This rules out the influences of the S-R instrumental pathway on action selection: rigid habits are not capable of driving the rat to make the suitable decision as they lack information on internal states (incidentally, note that this experimental condition was precisely designed by Balleine et al. (2003) to stop the effects of habits that would otherwise 'mask' the motivation-sensitive Pavlovian influence on action selection). On the other hand, satiation stops only one of the two influences of the Amg-NAc pathway on action selection in that it inhibits only the amygdala representation of the conditioned stimulus which has been satiated (compare the nac_{lev} activation in the two test phases). The fact that the Amg-NAc pathway 'votes' only for the action associated with the non-satiated food breaks the symmetry and makes the related action to win the competition in PM with a high chance (compare the pm_{lev} and m_{lev} activations in the two test phases).

The comparison between the lesioned and non-lesioned conditions (Fig. 2, last two blocks) reproduces the basic finding of the target experiment of Balleine et al. (2003) and confirms the aforementioned interpretation of the devaluation tests: as it happens in real rats, a lesion to the BLA pathway linking the amygdala to the NAc prevents the devaluation of food from having any effect on the action selection process. More in particular (see Fig. 3), during the test non-lesioned (Sham) rats perform the action associated with the non-devaluated food A 11.2 times on average whereas they perform the action associated with the devaluated food A 2.9 times on average: the difference between the two conditions is statistically significant (paired t-test, $t = 15.7003, df = 19, p < 0.001$). On the contrary, BLA-lesioned rats select actions randomly as indicated by

the fact that the number of performed actions associated with the non-devaluated and the devaluated food A have an average of 6.2 and 6.5: the difference between the two conditions is not statistically significant (paired t-test, $t = -0.4346, df = 19, p > 0.05$). These results show the plausibility of the hypothesis for which the Amg-(BLA)-NAc pathway bridges the Pavlovian processes happening in the amygdala with the instrumental processes happening in the cortex-basal ganglia pathway, so allowing the current state of animals' motivational system to modulate *on the fly* their action selection mechanisms.

7 Conclusions and Future Work

This paper presented an embodied model of some important relations existing between Pavlovian and instrumental conditioning. The model's architecture and functioning was constrained with relevant neuroscientific knowledge on the brain anatomy and physiology. The model was validated by successfully reproducing the primary outcomes of some instrumental conditioning devaluation tests conducted with normal and amygdala-lesioned rats. These tests are particularly important for studying the Pavlovian-instrumental interplay as they show how the sensitivity to motivational states exhibited by the Pavlovian system can transfer to instrumentally acquired behaviors.

To the best of the authors' knowledge, the model represents the first attempt to propose a comprehensive interpretation of the aforementioned phenomena, tested in an embodied model. The works most closely related to this one are those of Armony et al. (1997), Dayan and Balleine (2002), Morén and Balkenius (2000), and O'Reilly et al. (2007). The model presented here differs from these works in that it proposes an embodied model (absent in all mentioned researches), presents a fully developed model (Dayan and Balleine, 2002, presented only a 'sketched' model), and tackles the issue of the relations existing between Pavlovian and instrumental conditioning (Armony et al., 1997, Morén and Balkenius, 2000, and O'Reilly et al., 2007, focussed only on Pavlovian conditioning).

Notwithstanding the proposed model has these several strengths, it will be improved along many directions in future work. The first limit of the work is that the model was tested with an embodied system where input signals were heavily pre-processed before being fed into the model in the form of 'localistic representations' (one neuron-one object), and where actions could be specified at a rather abstract level by relying on hardwired low-level behavioral routines. In the future the whole model, or some of its parts (e.g. the amygdala component), will be tested with more challenging embodied systems where the model will be fed with realistic distributed input patterns (e.g., the activations of retina's pixels) and will be required to issue low-level motor commands (e.g., the desired displacement and turning speed). Second, the model has several limitations with respect to available biological evidence. For example, it does not learn to inhibit the dopamine signal at the onset of the USs if these are preceded by CSs, as it happens in real organisms (Schultz, 2002). This prevents the model from performing 'extinction' (i.e., to un-learn a classical conditioning association or an instrumental response if these are not followed anymore by a reward) and from stopping the weights' update. In future work, the model will be added this capability by drawing ideas from other works, for example O'Reilly et al. (2007). Moreover, the model cannot reproduce classical-conditioning based modulation of the *vigor* with which instrumental actions are performed (Niv et al., 2007), nor it is capable of triggering innate actions on the basis of classical-conditioning (e.g. approaching an US, or approaching a CS after this has been associated to an US; Dayan and Balleine, 2002). Finally, the model assumes that the selection of actions takes place within premotor cortex. However, there is strong evidence (Redgrave et al., 1999) that in real brains action selection takes place at the level of the DLS itself, and so PM activations might only reflect such selection without causing it (cf. Cisek, 2007). This possibility, however, opens up the problem of how the NAc might influence such action selection, as requested for the Pavlovian processes to exert an influence on instrumental processes. In this respect, an interesting neural pathway through which this influence might be implemented are the striato-nigro-striatal connections (or 'dopaminergic spirals'; Haber et al., 2000). These topics will be addressed in future work.

Notwithstanding these limitations, the proposed model represents an important step in the construction of an integrated picture on how animals' motivational systems can both drive instrumental learning and directly regulate behavior. Constructing such a picture is of paramount importance from the scientific point of view as psychology and neuroscience have now amassed a large body of evidence and knowledge on the phenomena investigated here which would greatly benefit of theoretical systematization. In this respect, we believe that computation modeling carried out under the principles of computational embodied neuroscience illustrated in Sect. 2 can greatly aid this process.

As mentioned in Sect. 1, although this papers has mainly a scientific relevance, the research agenda of the work presented here has also a potential interest for overcoming the limited autonomy of current robots. In fact, a way to tackle these limits is to attempt to understand the mechanisms underlying organisms' behavioural flexibility so as to use

them in designing robot's controllers. In this respect, notwithstanding the motivational and emotional regulation of behavior is very important for behavioural flexibility, it has been almost completely overlooked by autonomous robotics. For this reason Parisi (2004) has advocated the need of an 'Internal Robotics' research agenda dedicated to the study of these processes. In line with this, recently machine learning and robotics communities have been devoting increasing efforts to the study of autonomous learning by trying to improve the standard reinforcement learning algorithms mentioned in Sect. 1 on the basis of ideas coming from the study of real organisms (Zlatev and Balkenius, 2001; Weng et al., 2001). In this respect, the investigations on emotional regulation of learning and behaviour in animals, such as those reported here, are expected to produce important insights on possible new principles and techniques to be used to design more powerful learning algorithms exhibiting a degree of autonomy similar to that of real organisms (see Barto et al., 2004, and Schembri et al., 2007, for two examples of this).

Acknowledgements

This research was supported by the EU-funded Integrated Project *ICEA - Integrating Cognition, Emotion and Autonomy*, contract no. FP6-IST-027819-IP. A preliminary version of this work was published as Mannella et al. (2007).

References

Armony, J. L., Servan-Schreiber, D., Romanski, L. M., and LeDoux, D. J. J. E. (1997). Stimulus generalization of fear responses: effects of auditorycortexlesions in a computational model and in rats. *Cereb. Cortex*, 7(2):157–165.

Baldassarre, G. (2008). Self-organization as phase transition in decentralized groups of robots: a study based on Boltzmann entropy. In Mikhail, P., (Ed.), *Advances in Applied Self-Organizing Systems*, pages 127–146. Springer-Verlag, Berlin.

Balleine, B. W., Killcross, A. S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *J. Neurosci.*, 23(2):666–675.

Balleine, B. W. and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.*, 29(5):272–279.

Barto, A., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *International Conference on Developmental Learning (ICDL)*, LaJolla, CA.

Baxter, M. G. and Murray, E. A. (2002a). The amygdala and reward. *Nat. Rev. Neurosci.*, 3(7):563–573.

Baxter, M. G. and Murray, E. A. (2002b). The amygdala and reward. *Nature Rev. Neurosci.*, 3(7):563–573.

Blair, H. T., Sotres-Bayon, F., Moita, M. A. P., and Ledoux, J. E. (2005). The lateral amygdala processes the value of conditioned and unconditioned aversive stimuli. *Neuroscience*, 133(2):561–569.

Brody, C., Pouget, A., Shadlen, M., and Zador, A., (Eds.) (2004). *Abstracts of Papers Presented at the 2004 Meeting on Computational & System Neuroscience*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraulaz, G., and Bonabeau, E., (Eds.) (2001). *Self-organization in biological systems*. Princeton University Press, Princeton, NJ.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.*, 26(3):321–352.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 362(1485):1585–1599.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. The MIT Press, Cambridge, MA.

Darwin, C. (1859). *The Origin of Species*. http://www.literature.org/authors/darwin-charles/the-origin-of-species/index.html.

Dayan, P. and Balleine, B. (2002). Reward, motivation and reinforcement learning. *Neuron*, 36:285–298.

Domjan, M. (2006). *Principles of Learning and Behaviour*. Thomson Wadsworth, Belmont, CA.

Haber, S. N., Fudge, J. L., and McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.*, 20(6):2369–2382.

Holland, O. and McFarland, D. (2001). *Artificial Ethology*. Oxford, Oxford University Press.

Houk, J. C., Adams, J. L., and Andrew, G. B. (1995). A model of how the basal ganglia generate and use neural signals ghat predict reinforcement. In Houk, J. C., Davids, J. L., and Beiser, D. G.,

(Eds.), *Models of Information Processing in the Basal Ganglia*, pages 249–270. The MIT Press, Cambridge, MA.

Knight, D. C., Nguyen, H. T., and Bandettini, P. A. (2005). The role of the human amygdala in the production of conditioned fear responses. *Neuroimage*, 26(4):1193–1200.

Kobayashi, Y. and Okada, K.-I. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. *Ann. N. Y. Acad. Sci.*

Langton, C., (Ed.) (1987). *The First International Conference on the Simulation and Synthesis of Living Systems (ALifeI)*.

Lieberman, D. A. (1993). *Behavior and Cognition*. Brooks/Cole, Pacific Grove, CA.

Mannella, F., Mirolli, M., and Baldassarre, G. (2007). The role of amygdala in devaluation: a model tested with a simulated robot. In Berthouze, L., Prince, C. G., Littman, M., Kozima, H., and Balkenius, C., (Eds.), *Proceedings of the Seventh International Conference on Epigenetic Robotics*, pages 77–84. Lund University Cognitive Studies.

Mannella, F., Zappacosta, S., and Baldassarre, G. (2008). A computational model of the amygdala nuclei's role in second order conditioning. In Tani, M. A. J., Hallam, J., and Meyer, J.-A., (Eds.), *Proceedings of the Tenth International Conference on Simulation of Adaptive Behavior: From Amimals to Animals 10*.

Maren, S. (2005). Building and burying fear memories in the brain. *Neuroscientist*, 11(1):89–99.

McDonald, A. J. (1998). Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.*, 55(3):257–332.

Meyer, J.-A. and Wilson, S. W., (Eds.) (1991). *From Animals to Animats 1: Proceedings of the First International Conference on Simulation of Adaptive Behaviour*. MIT Press, Cambridge, MA.

Mogenson, G. J., Jones, D. L., and Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog. Neurobiol.*, 14(2-3):69–97.

Morén, J. and Balkenius, C. (2000). A computational model of emotional learning in the amygdala. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H. L., and Wilson, S. W., (Eds.), *From Animals to Animats 6: Proceedings of the 6th International Conference on the Simulation of Adaptive Behaviour*, Cambridge, Mass. The MIT Press.

Newell, A. (1973). You can't play 20 questions with nature and win: projective comments on pthe papers of this symposium. *Visual Information Processing*, pages 135–183.

Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *J. Psychopharmacol.*, 191(3):507–520.

Nolfi, S. (2006). Behaviour as a complex adaptive system: on the role of self-organization in the development of individual and collective behaviour. *ComplexUs*, 2(3-4):195–203.

Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics: The Biology,Intelligence,and Technology*. MIT Press, Cambridge, MA, USA.

O'Reilly, R., Frank, M., Hazy, T., and Watz, B. (2007). PVLV: The primary value and learned value pavlovian learning algorithm. *Behav. Neurosci.*, 121:31–49.

Packard, M. G. and Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.*, 25:563–593.

Parisi, D. (2004). Internal robotics. *Connection Science*, 16(4):325–338.

Parisi, D., Cecconi, F., and Nolfi, S. (1990). Econets: Neural networks that learn in an environment. *Network*, 1:149–168.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press., London.

Pitkänen, A., Jolkkonen, E., and Kemppainen, S. (2000). Anatomic heterogeneity of the rat amygdaloid complex. *Folia Morphol. (Warsz)*, 59(1):1–23.

Prescott, T. J., Gonzalez, F. M., Humphries, M., and Gurney, K. (2003). Towards a methodology for embodied computational neuroscience. In *Proceedings of the Symposium on Scientific Methods for the Analysis of Agent-Environment Interaction (AISB2003)*. AISB Press.

Prescott, T. J., Gonzlez, F. M. M., Gurney, K., Humphries, M. D., and Redgrave, P. (2006). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Netw*, 19(1):31–61.

Price, J. L. (2003). Comparative aspects of amygdala connectivity. *Ann. N.Y. Acad. Sci.*, 985(1):50–58.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *J. Neurosci.*, 89(4):1009–1023.

Rolls, E. T. (2005). Taste and related systems in primates including humans. *Chem. Senses*, 30 Suppl 1:i76–i77.

Schembri, M., Mirolli, M., and Baldassarre, G. (2007). Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In Demiris, Y., Mareschal, D., Scassellati, B., and Weng, J., (Eds.), *Proceedings of the 6th International Conference on Development and Learning (ICDL)*, pages E1–6, London. Imperial College.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36:241–263.

Sejnowski, T. J., Koch, C., and Churchland, P. S. (1988). Computational neuroscience. *Science*, 241(4871):1299–1306.

Shi, C. and Davis, M. (1999). Pain pathways involved in fear conditioning measured with fear-potentiated startle: lesion studies. *J. Neurosci.*, 19(1):420–430.

Skinner, B. (1938). *The Behavior of Organisms*. Appleton Century Crofs, New York, NY.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Sutton, R. S. and Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychol. Rev.*, 88:135–140.

Thorndike, E. L. (1911). *Animal Intelligence*. Transaction Publishers, Rutgers, NJ.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291:599–600.

Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Rev. Neurosci.*, 7:464–476.

Zlatev, J. and Balkenius, C. (2001). Introduction: Why epigenetic robotics? In Balkenius, C., Zlatev, J., Kozima, H., , Dautenhahn, K., and Breazeal, C., (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, volume 85 of *Lund University Cognitive Studies*, pages 1–4.